Large Language Models Empowered Autonomous Edge AI for Connected Intelligence

Abstract

The evolution of wireless networks gravitates toward connected intelligence, a concept that envisions seamless interconnectivity among humans, objects, and intelligence in a hyper-connected cyber-physical world. Edge artificial intelligence (Edge AI) is a promising solution to achieve connected intelligence by delivering high-quality, low-latency, and privacy-preserving AI services at the network edge. This article presents a vision of autonomous edge AI systems that automatically organize, adapt, and optimize themselves to meet users' diverse requirements, leveraging the power of large language models (LLMs), that is, generative pretrained transformer (GPT). By exploiting the powerful abilities of GPT in language understanding, planning, and code generation, as well as incorporating classic wisdom such as task-oriented communication and edge federated learning, we present a versatile framework that efficiently coordinates edge AI models to cater to users' personal demands while automatically generating code to train new models in a privacy-preserving manner. Experimental results demonstrate the system's remarkable ability to accurately comprehend user demands, efficiently execute AI models with minimal cost, and effectively create high-performance AI models at edge servers.